

Michael Hässig

Ungenauere Erkenntnis? Die Bedeutung von Wahrscheinlichkeit für die medizinische Praxis

Jeder Arzt macht es: Wenn aufgrund der Anamnese, der Vorgeschichte, der zeitlichen und örtlichen Vorgaben und der Symptomatik die Wahrscheinlichkeit für eine bestimmte Krankheit spricht, so nennt dies der Arzt Diagnose und handelt danach. Dies ist nicht verwerflich, sondern erhält die medizinische Handlungsfähigkeit bei unvollkommener Informationslage, wie sie fast immer in der Klinik vorliegt. Dieses Vorgehen leitet auch den Laien, ohne dass er sich der theoretischen Grundlagen seines Handelns bewusst wird.

Im klinischen Setting muss der Arzt eine Risikoeinschätzung vornehmen. Meist macht er dies mehr oder weniger willkürlich, da eine genauere Abklärung aus zeitlichen Gründen oft gar nicht möglich ist oder auch keine einschlägigen Daten zur Verfügung stehen. Viele Studien zeigen, dass Mediziner oft Schwierigkeiten mit der Risikoeinschätzung haben und das Risiko oft überschätzen.¹

Worum es mir im Folgenden geht, sind keine wissenschaftstheoretischen Fragen, sondern die intuitive und evidenzbasierter Arbeit des klinischen Praktikers. Zunächst soll gezeigt werden, wie die Risikoperzeption unserer subjektiven Wahrnehmung unterliegt und inwiefern philosophische und erkenntnistheoretische Konzepte dazu beigetragen haben, unsere Wahrnehmung zu objektivieren. 1987 hat Paul Slovic in einer vielbeachteten Arbeit

¹ Odette Wegwarth, Gerd Gigerenzer: Vom Risiko, Risiken zu kommunizieren, in: *Therapeutische Umschau* 64 (2007) 687–692.

aufgezeigt, wie unterschiedlich verschiedene Gruppierungen eine Risikoeinschätzung vornehmen (Abb. 1). Die Einschätzung ist neben der klassischen Formel «Risiko = Schaden x Eintretenswahrscheinlichkeit» auch von Ort und Zeit abhängig. Die Ungenauigkeit der Schätzung nimmt zu, wenn ein Ereignis sehr selten eintritt, aber einen sehr grossen Schaden verursacht oder umgekehrt.

Activity or technology	League of Women Voters	College students	Active club members	Experts
Nuclear power	1	1	8	20
Motor vehicles	2	5	3	1
Handguns	3	2	1	4
Smoking	4	3	4	2
Motorcycles	5	6	2	6
Alcoholic beverages	6	7	5	3
General (private) aviation	7	15	11	12
Police work	8	8	7	17
Pesticides	9	4	15	8
Surgery	10	11	9	5
Fire fighting	11	10	6	18
Large construction	12	14	13	13
Hunting	13	18	10	23
Spray cans	14	13	23	26
Mountain climbing	15	22	12	29
Bicycles	16	24	14	15
Commercial aviation	17	16	18	16
Electric power (non-nuclear)	18	19	19	9
Swimming	19	30	17	10
Contraceptives	20	9	22	11
Skiing	21	25	16	30
X-rays	22	17	24	7
High school and college football	23	26	21	27
Railroads	24	23	29	19
Food preservatives	25	12	28	14
Food coloring	26	20	30	21
Power mowers	27	28	25	28
Prescription antibiotics	28	21	26	24
Home appliances	29	27	27	22
Vaccinations	30	29	29	25

Abb. 1: Paul Slovic: Perception of Risk, in: Science 236 (1987) 280–285.

Um die einzelnen Fakten wie Falldefinition, Anamnese, Symptomatik miteinander verknüpfen zu können, stehen verschiedene epidemiologische Modelle zur Verfügung. Eines davon ist das Bayes-Theorem, das in der klinischen Epidemiologie breite Anwendung findet. Das Bayes-Theorem stellt einen konzeptionellen, quantitativen Ansatz zur Erfassung klinischen Handelns dar.² Es basiert auf dem erkenntnistheoretischen Konzept der Induktion und wurde von Austin Bradford Hill für die Medizin nutzbar gemacht.³

Die moderne Erkenntnistheorie geht nicht mehr davon aus, dass wir in der Lage sind, die Dinge objektiv zu erkennen. Der deduktive Ansatz Karl Poppers war der letzte «objektivistische» Versuch, Kausalität zu begründen. Popper hält fest, dass eine induktive Objektivität nicht möglich ist und versucht, Objektivität deduktiv zu retten. «Jede Hypothese ist richtig, bis das Gegenteil bewiesen ist», das heisst – in die medizinische Praxis übersetzt – jede Behandlung ist solange gut, bis ihre schädlichen Nebenwirkungen bekannt sind – hier zeigen sich die Grenzen von Poppers Theorie im medizinischen Setting.⁴ Der Mediziner kann nicht beliebig testen. Er ist den Patienten verpflichtet. Er ist sich aber durchaus bewusst, dass er Entscheide treffen muss mit einer unsicheren Datenlage, in der Hoffnung, dass seine Entscheide richtig sind und dem Patienten helfen.

Klinisches Handeln ist immer einer gewissen Ungenauigkeit unterworfen und beruht letztlich auf Wahrscheinlichkeiten. Sehr vereinfacht lautet das Bayes-Theorem: «nachgängige Wahrscheinlichkeit = vorgängige Wahrscheinlichkeit \times Testrate (positiv oder negativ).

² Ian Stewart: Mathematische Unterhaltung, in: Spektrum der Wissenschaft 7 (1997) 8–10; Roy Matthews: How right can you be?, in: New Scientist 2072 (1997) 28–33.

³ Austin Bradford Hill: The environment and disease: Association or causation?, in: Proceedings of the Royal Society of Medicine 58 (1965) 295–300.

⁴ Alfredo Morabia: On the origin of Hill's causal criteria, in: Epidemiology 2 (1991) 367–369.

Aufgrund eines bestimmten Symptoms besetzt für den Patienten eine Wahrscheinlichkeit P , die vorgängige Wahrscheinlichkeit, (anterior probability), dass das Symptom durch eine bestimmte Krankheit verursacht worden ist. Ist diese Wahrscheinlichkeit genügend hoch, das heisst, überschreitet sie eine gewisse Schwelle (threshold level), wird der Arzt keine weiteren Abklärungen vornehmen. Er wird das Resultat einer Diagnose zuordnen und gegebenenfalls handeln und eine Therapie vorschlagen. Erreicht die vorgängige Wahrscheinlichkeit diesen Schwellenwert, der zum Handeln veranlasst, nicht, werden weitere Untersuchungen vorgenommen mit dem Ziel, die Wahrscheinlichkeit über oder gezielt unter diesen Schwellenwert zu bringen. Die Wahrscheinlichkeit nach dem Test ist die nachgängige Wahrscheinlichkeit (posterior probability). Inwieweit eine diagnostische Abklärung die nachgängige Wahrscheinlichkeit beeinflusst, hängt ab von der Sensitivität (Anteil erkrankter Individuen mit richtig positivem Testresultat gegenüber allen Erkrankten) und Spezifität (Anteil gesunder Individuen mit richtig negativem Testresultat gegenüber allen Gesunden) des Testsystems sowie der Prävalenz der zu untersuchenden Krankheit.

Der Grenzwert, der eine Diagnose zulässt oder zum Handeln veranlasst, wird vom Arzt aufgrund vorgängig definierter Kriterien festgelegt. Man geht davon aus, dass jede Manipulation am Patienten, ob Test oder Therapie, auch ein absolutes Risiko beinhaltet. Für den Grenzwert zum Handeln wird die Summe aus Test- und Behandlungsrisiko dem Behandlungserfolg gegenübergestellt. Diese Gegenüberstellung wird als Qualität der Behandlung bezeichnet. Zur Qualitätsberechnung können diverse Parameter herangezogen werden wie beispielsweise Überlebenszeit, qualitätsadjustierte Überlebenstage, Leistungsmerkmale wie auch der finanzielle Vorteil einer Behandlung. Im Unterschied zur Humanmedizin können solche Grenzwerte in der Nutztiermedizin einfacher festgelegt werden, da die ethische Komponente anders gewichtet wird und die wirtschaftliche Komponente oft eine wesentliche Rolle spielt.

Die Wahrscheinlichkeitsrate «positiver Test» kann als Wert bezeichnet werden, mit dem die nachgängige Wahrscheinlichkeit steigt, wenn der Test positiv ausfällt. Sinngemäss gilt das Gleiche für ein negatives Testresultat. Die Wahrscheinlichkeitsrate «negativer Test» kann als Wert bezeichnet werden, mit dem die Wahrscheinlichkeit sinkt, wenn der Test negativ ausfällt.⁵

Jedes biologische Testsystem beinhaltet eine gewisse Ungenauigkeit. Bei kontinuierlich verlaufenden Testresultaten besteht in der Regel eine unexakte Trennschärfe. Bei konkreten Zuständen, zum Beispiel gesund/krank, ergibt sich daher fast immer ein Anteil, der aufgrund des Testresultats nicht korrekt zugeordnet wird (falsch positiv oder falsch negativ). Die Qualität eines Tests wird primär festgehalten durch die Sensitivität und Spezifität als Mass der Trennschärfe zwischen zwei Zuständen, die normalerweise mit positiv und negativ charakterisiert werden. Um Aussagen über die Qualität eines Tests machen zu können, muss der wirkliche Zustand des Testobjekts bekannt sein. Oft wird ein neuer Test mit einem Standardtest – goldener Standard – oder ein Schnelltest mit einem komplizierten Labortest verglichen. Es ist allerdings schwierig, einen goldenen Standard festzulegen. In der Regel wird derjenige Test zum goldenen Standard erklärt, der bereits etabliert ist oder die höchste Sensitivität und Spezifität aufweist.

Da kein Testsystem mit absoluter Sicherheit zwischen zwei Zuständen unterscheiden kann, stellt das Festlegen eines Grenzwertes und somit das Festlegen der Sensitivität und Spezifität immer eine Optimierungsaufgabe dar. Diesem Umstand wird in epidemiologischen Studien bei Überwachungstests (surveillance) oder Vorsorgetests (screening) Rechnung getragen, indem zwei unabhängige Testsysteme verwendet werden, wobei der eine Test

⁵ Roman Jaeschke, Gordon H. Guyatt, David L. Sackett: Users' guide to Medical Literature; III. How to Use an Article About a Diagnostic test; B. What Are the Results and Will They Help Me in Carrying for My Patients?, in: The Journal of the American Medical Association 271 (1994) 703–707.

eine hohe Sensitivität und der andere Test eine hohe Spezifität aufweist. Dies ist vor allem dann wichtig, wenn das Resultat eines Tests mit schwerwiegenden Konsequenzen verbunden ist.

Die Wahrscheinlichkeitsrate für einen positiven beziehungsweise negativen Test lässt sich als Funktion aufzeichnen (Abb. 2).

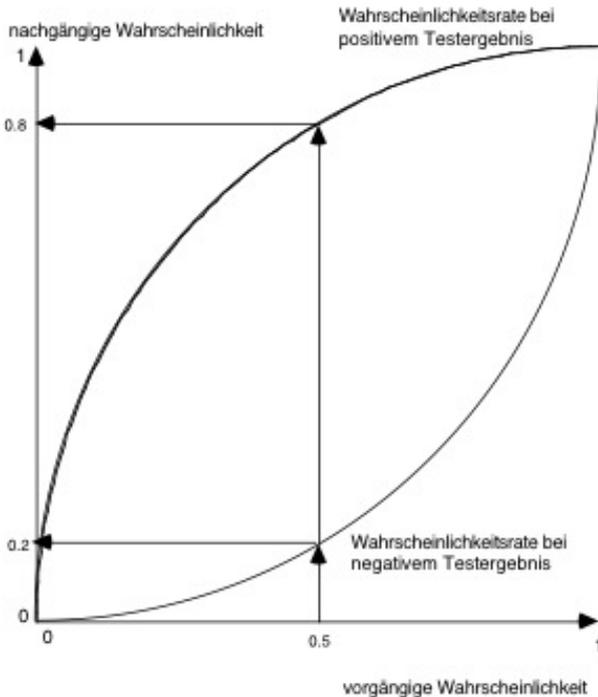


Abb. 2: Vorgängige und nachgängige Wahrscheinlichkeit bei gegebener Sensitivität (90%) und Spezifität (80%) mit dem Beispiel einer vorgängigen Wahrscheinlichkeit von 50% ($P = 0.5$) und den jeweils resultierenden nachgängigen Wahrscheinlichkeiten bei positivem beziehungsweise negativem Testergebnis.⁶

⁶ Jacobus Lubsen: Vorlesungsunterlagen in klinischer Epidemiologie im Rahmen des MPH Curriculums der Deutschschweizer Universitäten (1995); Michael Hässig: Vorgehen in Betrieben mit gehäuften Verwerfen beim Rind, Habilitationsschrift Universität Zürich (1999).

Aus Abb. 2 lässt sich auch ableiten, dass ein Test seine optimale Aussage bei einer vorgängigen Wahrscheinlichkeit von $P = 0.5$ hat. Bei offensichtlichem Vorliegen einer Krankheit, das heisst bei einer vorgängigen Wahrscheinlichkeit über 0.9, oder bei seltenen Krankheiten, das heisst bei einer vorgängigen Wahrscheinlichkeit unter 0.1, hat jeder Test eine verminderte Aussagekraft.

Der Grenzwert (threshold), der eine Diagnose zulässt beziehungsweise zum Handeln veranlasst, muss vom Untersuchenden festgelegt werden. Können Aufwand und Ertrag einer Massnahme objektiviert werden, so lässt sich auch der sinnvolle Testbereich einschränken und der Grenzwert für Handlungsbedarf festlegen. In der klinischen Epidemiologie gilt der Grundsatz, dass oberhalb des Grenzwertes für Handlungsbedarf eine Behandlung vorgenommen wird (intention to treat). Man geht davon aus, dass jede Manipulation am Patienten, ob Test oder Therapie, auch ein absolutes Risiko (Inzidenz) für einen unerwünschten Ausgang beinhaltet, auch wenn dieser oft sehr klein ist.

Wenn ein Individuum gesund ist und keine Therapie durchgeführt wird, hat es die grösste Überlebenschance (1.0). Ist das Individuum gesund und die Therapie wird durchgeführt, so vermindert sich die Überlebenschance um das Risiko, das von der Therapie ausgeht. Das Risiko der Therapie sind die unerwünschten Nebenwirkungen. Werden kranke Individuen behandelt, so überlebt im Verhältnis zu den gesunden, unbehandelten ein gewisser Anteil. Da fast jede Krankheit auch eine gewisse Selbstheilungsrate aufweist, wird auch ohne Therapie ein gewisser Anteil überleben.

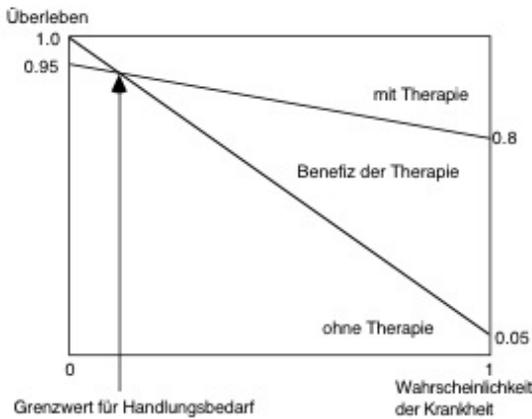


Abb. 3: Festlegung des Grenzwertes für Handlungsbedarf nach Abschätzung des Risikos durch die Massnahme

Mithilfe des Bayes-Theorems und der festgelegten Grenzwerte für Massnahmen lassen sich Entscheidungshilfen für den Kliniker erstellen. Klinisches Handeln erfolgt oft unbewusst in Anlehnung an dieses Modell (Abb. 4).

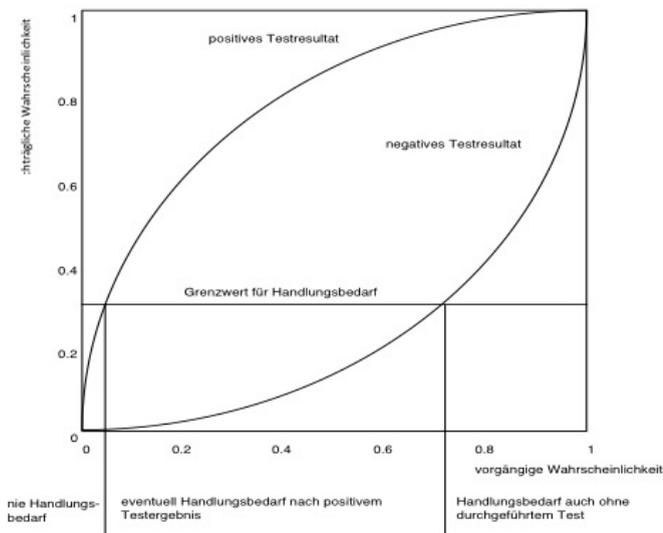


Abb. 4: Das quantitative, konzeptionelle Modell des Bayes-Theorems. Aus: M. Hässig: Vorgehen in Betrieben mit gehäuften Verwerfen beim Rind.

Abb. 4 unterscheidet sich von Abb. 2 lediglich dadurch, dass Grenzwerte eingeführt wurden. In der nachgängigen Wahrscheinlichkeit gibt es einen Grenzwert, der zum Handeln beziehungsweise zu Massnahmen Anlass gibt. In der vorgängigen Wahrscheinlichkeit gibt es zwei Grenzwerte. Diese halten fest, in welchem Bereich der vorgängigen Wahrscheinlichkeit ein Test überhaupt sinnvoll erscheint. Unterhalb des unteren Grenzwertes beeinflusst der Ausgang des Tests die Entscheidung zum Handeln nicht, da weder bei positivem noch negativem Testausgang die nachgängige Wahrscheinlichkeit oberhalb des Grenzwertes Handlungsbedarf vorliegt. Oberhalb des oberen Grenzwertes beeinflusst der Ausgang des Tests die Entscheidung zum Handeln auch nicht, da weder bei positivem noch negativem Testausgang die nachgängige Wahrscheinlichkeit unterhalb des Grenzwertes ein Handeln erfordert. Die Grenzwerte, bei denen kein Handlungsbedarf (test threshold) beziehungsweise in jedem Fall ein Handlungsbedarf (treatment threshold) besteht, können berechnet werden.⁷

Nun muss in einem nächsten Schritt die Frage beantwortet werden: Wenn ein Resultat kausal ist, wann ist es dann evident? Als evidenzbasierte Medizin wird nun die strukturierte Präsentation der medizinischen Resultate bezeichnet. Diese Definition fokussiert auf die rein wissenschaftliche Beurteilung klinischer Resultate. Die Betreuung des Patienten steht hier nicht im Mittelpunkt. Daher sehen andere Definitionen die Präsentation von Ergebnissen nicht im Mittelpunkt, sondern die patientenorientierte Entscheidung auf der Grundlage von empirischen Daten.

Robert Koch stellte bereits 1882 seine Postulate zur Kausalität auf, die für ihn hinreichend zur Erklärung eines Infekts waren:

- Der Erreger muss in jedem Fall von Krankheit nachweisbar sein.
- Er muss kultivierbar sein.
- Er muss die Krankheit auslösen.
- Er muss rekultivierbar sein.
- Er darf bei keiner anderen Krankheit vorkommen.

⁷ Ronald D. Smith: Veterinary Clinical Epidemiology, Boca Raton etc. ²1995.

Schon bald stellte sich heraus, dass diese Postulate nicht umfassend sein konnten, weil sie die Umwelteinflüsse und weitere Faktoren nicht berücksichtigten. Erst 1965 stellte Austin Bradford Hill, neue allgemein anerkannte Kausalitätskriterien auf:

- Gesicherte Assoziation: Der Zusammenhang muss statistisch genügend gesichert sein.
- Konstante Assoziation: Die Resultate müssen unter verschiedenen Umständen gleichbleiben.
- Zeitliche Chronologie: Die Ursache kommt vor der Krankheit.
- Dosis-Wirkung-Zusammenhang
- Biologische Plausibilität: Der Zusammenhang muss nach heutigem Wissensstand erklärbar sein.
- Experimentelle Bestätigung
- Analogie: Ähnliche Ursachen müssen zu ähnlichen Krankheiten führen.
- Kohärenz: Die Resultate müssen reproduzierbar sein.

Auch die Kausalitätskriterien von Bradford Hill sind nicht unumstritten. Der Dosis-Wirkung-Zusammenhang ist oft nicht gegeben. Hormone haben ihr Wirkungsoptimum. Toxische hormonähnliche Substanzen wie Biphenol A sind in geringen Mengen toxischer. Die biologische Plausibilität – Der Zusammenhang muss nach heutigem Wissensstand erklärbar sein – kann ein unzuweckmässiges Kriterium sein. Sie kann den wissenschaftlichen Fortschritt und die Erkenntnis behindern. So vergingen rund fünfzehn Jahre bis akzeptiert wurde, dass ein infektiöses Agens nicht unbedingt Erbsubstanz zu enthalten hat. Dies stand der Prionentheorie zu spongiformer boviner Enzephalopathie (BSE) im Wege. Sehr oft werden Assoziationen, die statistisch gut belegt werden können, und Kausalität gleichgesetzt. Dies ist ein grosser Trugschluss unserer Tage. Kausalität besteht aus Assoziationen und weiteren auf der Logik beruhenden Postulaten.⁸

⁸ Robert Koch: The etiology of tuberculosis [Die Ätiologie der Tuberkulose, in: Berliner Klinische Wochenschrift 15 (1882) 221–230, translated and edited by Thomas D. Broke], in: Reviews of Infectious Diseases 4 (1982)

Die zukünftigen klinischen Fragen müssen sich daher mit der Frage auseinandersetzen, welchen Beitrag Wahrscheinlichkeiten zur Gültigkeit – und nicht zur Objektivität – medizinischer Diagnosen leisten. Denn der Kliniker, will er handlungsfähig bleiben, ist verpflichtet, zu handeln: auf der Grundlage des klinischen Axioms der «intention to treat» und des aktuellen Wissensstands. Für das klinische Handeln ist sind ungenaue Erkenntnis und Wahrscheinlichkeit grundlegende Elemente einer best practice.

conexus 2 (2019) 52–62

© 2019 Michael Hässig. Dieser Beitrag darf im Rahmen der Lizenz CC BY-NC-ND 4.0 – Creative Commons: Namensnennung/nicht kommerziell/keine Bearbeitungen – weiterverbreitet werden.



<https://doi.org/10.24445/conexus.2019.02.005>

Prof. Dr. Michael Hässig, Universität Zürich, Vetsuisse-Fakultät, Abteilung Agrovet-Strickhof, Eschikon 27, 8315 Lindau
mhaessig@vetclinics.uzh.ch

1270–1274; A. B. Hill: The environment and disease; A. Morabia: On the origin of Hill's causal criteria.